

On Terms

Observer Drift: A Drifting Definition

G. A. Smith

University of Delaware

Multiple meanings for a single descriptive term can contribute to conceptual confusion and a general inability to distinguish reliably among different phenomena. A particular case in point is "observer drift." The current use of three definitions for this term may lead to various problems ranging from simple, inconsistent interpretations of the phenomenon, and the evolution of new and perhaps unnecessary terms, to more complex problems such as the ineffective training of observers, improper application of countermeasures to attenuate observer drift, and failures to replicate studies concerning this and related phenomena. The purpose of this brief note is to describe these inconsistencies and suggest some possible alternative uses.

First, the term "observer drift" has been used, by and large, to describe the implicit changes in code definitions made by observers over time (e.g., Kazdin, 1977). In some instances, the meaning of the term has quite reasonably been extended to include the development of idiosyncratic versions of explicit definitions when observers work in pairs (e.g., Johnson & Bolstad, 1973); this has been referred to as "consensual observer drift."

Second, "observer drift" has also been used to describe observer reactivity to monitoring (e.g., DeMaster, Reid, & Twentyman, 1977; Taplin & Reid, 1973). Research in this area has shown that observer awareness of monitoring for interobserver agreement results in differential performances by observers.

Specifically, on days when observers know that monitoring will occur, agreement scores are markedly higher.

Third, the problem of definitional drift has been compounded further by the use of the term "observer drift" as the substantial drop in interobserver agreement scores coinciding with the shift from the end of training to the beginning of data collection (e.g., DeMaster et al., 1977; Taplin & Reid, 1973). Wildman and Erickson (1977) appear to be the first to have noted the expanded use of the term in this sense and, in the interest of clarity, referred to the "drop" phenomenon described by Taplin and Reid (1973) as "reactive effects of testing" (p. 261).

This latter attempt to reduce confusion by developing new terms to distinguish other phenomena from observer drift has been demonstrated by other authors as well. Weinrott and Jones (1984), for example, referred to observer reactivity to monitoring as "covert decline"; Smith and Sheaffer (1984) regarded the decrease in observer accuracy when shifting from training to formal data collection as "observer drop."

Others have used "observer drift" as a general label for drift, drop, and decline phenomena (e.g., DeMaster et al., 1977; Hartmann & Wood, 1982) or have cited drop and decline studies as evidence for drift (e.g., Kazdin, 1977). This practice not only perpetuates the confusion but also demonstrates that true observer drift (definitional drift) is not limited to observers involved in data collection—the phenomenon appears to affect researchers as well.

Recognizing that a problem exists is often the first step towards its solution. In the case of "observer drift," awareness of (1) reactivity ("decline") and (2) "drop" as separate phenomena may eventually

The author gratefully acknowledges the assistance of George Brabner, Alan Kazdin, and an anonymous reviewer for comments on a draft of this article. Reprints may be obtained from the author, Educational Studies, University of Delaware, Newark, DE 19716.

lead to clearer definitions and reduce the probability of observer inconsistency. Resolving definitional problems by generating new terms, however, may create new problems such as the acceptance of a new term without careful consideration. For example, "drop" and "decline" only cover decreases in observer agreement scores, unlike "observer reactivity," which covers increases as well as decreases.

In conclusion, "observer drift" should be limited to gradual definitional drift as originally defined. The use of simple language and distinct terms should attenuate confusion and prevent problems in replicating or applying research on observer performance. In general, the confusion of terminology discussed in this note points out the need for investigating variables controlling the verbal behavior of researchers.

REFERENCES

- DeMaster, B., Reid, J., & Twentyman, C. (1977). The effects of different amounts of feedback on observer's reliability. *Behavior Therapy, 8*, 317-329.
- Hartmann, D., & Wood, D. (1982). Observational methods. In A. Bellak, M. Hersen, & A. Kazdin (Eds.), *International handbook of behavior modification and therapy* (pp. 109-138). New York: Plenum.
- Johnson, S. M., & Bolstad, O. D. (1973). Methodological issues in naturalistic observation: Some problems and solutions for field research. In L. A. Hamerlynck, L. C. Handy, & E. G. Mash (Eds.), *New developments in behavioral research: Theory, methods, and applications* (pp. 167-188). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kazdin, A. E. (1977). Artifact, bias, and complexity of assessment: The ABCs of reliability. *Journal of Applied Behavior Analysis, 10*, 141-150.
- Smith, G. A., & Sheaffer, B. (1984). Observer reactivity in monitored and unmonitored analogue conditions. *Journal of Psychoeducational Assessment, 2*, 249-255.
- Taplin, P. S., & Reid, J. B. (1973). Effects of instructional set and experimenter influence on observer reliability. *Child Development, 44*, 547-554.
- Weinrott, M., & Jones, R. (1984). Overt versus covert assessment of observer reliability. *Child Development, 55*, 1125-1137.
- Wildman, B. G., & Erickson, M. T. (1977). Methodological problems in behavioral observation. In J. Cone & R. Hawkins (Eds.), *Behavioral assessment* (pp. 255-273). New York: Bruner/Mazel Publishers.